# EXECUTIVE SUMMARY

This addendum to the draft Background Review Documents (BRDs) on four *in vitro* test methods – the Isolated Rabbit Eye (IRE) assay, the Isolated Chicken Eye (ICE) assay, the Bovine Corneal Opacity and Permeability (BCOP) assay, and the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assay – for detecting ocular corrosives and severe irritants (Available: http://iccvam.niehs.nih.giv/methods/ocudocs/ocu_brd.htm [NICEATM 2004]) contains the results of the accuracy and reliability reassessment conducted on each of the four test methods (Available: http://iccvam.niehs.nih.gov/methods/ocudocs/reanalysis.htm [NICEATM 2005b]). This reassessment was in response to:

- the submission of additional *in vitro* test data and/or corresponding *in vivo* rabbit eye test data provided to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in response to a second *Federal Register* (*FR)* notice (Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [NIEHS 2005] requesting all available *in vitro* data on these four *in vitro* ocular irritancy test methods and corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either via ethical human studies or accidental exposure)
- clarification of the European Union (EU) (EU 2001) and United Nations (UN) Globally Harmonized System (GHS) (UN 2003) ocular hazard classification rules for severe irritants (Available: http:/www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html); this resulted in the reclassification of 10 to 15 substances from nonsevere to severe irritants, depending on the *in vitro* ocular irritancy test method and the ocular hazard classification system used
- the reassignment of substances to chemical classes using Medical Subject Headings (MeSH) (Available: http://www.nlm.nih.gov/mesh [NLM 2005]), an internationally recognized standardized classification system that would ensure consistency in classifying substances by chemical class
- a recommendation that the accuracy analysis consider whether a substance was classified as corrosive or severely irritating based on the severity of the response and/or its persistence to day 21 post-treatment

A list of proposed reference substances for validation of *in vitro* tests to detect ocular corrosives and severe irritants was included in the draft BRDs released on November 1, 2004 [NICEATM 2004]. This addendum provides a revised list of proposed reference chemicals, which was prepared after consideration of the following:

- recommendations of the Expert Panel that resulted from their deliberations on January 11-12, 2005 (Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [NICEATM 2005a])
- clarification regarding the GHS rules for classification of severe irritants [UN 2003] that resulted in the reclassification of two proposed reference substances from nonsevere to severe irritants
- reassignment of the candidate reference substances to chemical classes using MeSH [NLM 2005]

- submission of additional Draize rabbit eye test results for approximately 300 substances

**Table ES-1** provides a comparison of the accuracy statistics for each *in vitro* test method re-evaluated in this addendum, when results are compared to the GHS ocular hazard classification system.

***IRE Test Method***

The IRE test method was developed by Burton et al. (1981) and proposed as a preliminary *in vitro* screen for the assessment of severe eye irritants. This organotypic test method is also referred to as the Rabbit Enucleated Eye Test (REET) (e.g., Guerriero et al. [2004]). The principal advantage of the IRE test is that it eliminates the use of live animals for ocular irritancy testing and thus the pain and suffering potentially associated with the *in vivo* Draize rabbit eye test. Another advantage of the IRE test method is that it typically uses eyes isolated from euthanized rabbits used for other research purposes or from animals sacrificed commercially as a food source. In the IRE, liquid or solid substances are placed directly on the corneal surface of isolated rabbit eyes, which are held and maintained in a temperature-controlled chamber. After a 10-second exposure, followed by rinsing, the treated eye is evaluated for corneal opacity, corneal swelling, fluorescein penetration, and effects on the corneal epithelium at various times over a four-hour observation period. Substances that exceed a defined cut-off value for any single one of these endpoints are then identified as corrosives or severe irritants.

No additional data were submitted for the IRE test method. The existing database of substances tested using the four ocular endpoints recommended in the draft IRE BRD (corneal opacity, corneal swelling, fluorescein penetration, and epithelial integrity) remained limited to the Guerriero et al. (2004) study. As recommended by the Expert Panel, a reanalysis was performed in which substances in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies that had been identified as ocular corrosives/severe irritants using appropriate decision criteria (i.e., a corneal opacity score greater than or equal to 3, or a corneal swelling equal to or greater than a 25%) were considered together with the test results obtained by Guerriero et al. (2004). This database is referred to as the "Expanded Data Set."

Substances that were identified as ocular corrosives/severe irritants based on *in vitro* results by any single endpoint were, therefore, included in the reanalysis as part of the expanded data set. Substances in the CEC (1991), Balls et al. (1995), and Gettings et al. (1996) studies that were identified as nonsevere irritants or nonirritants, based on *in vitro* results, were not included in the expanded data set. These substances were not included because an evaluation that included any of the omitted endpoints might have resulted in a severe irritant classification. For example, a substance that did not produce ≥ 25% corneal swelling might have produced a corneal opacity score, fluorescein penetration score, or damage to the epithelium that would have classified it as a severe irritant had these endpoints been evaluated.

**Table ES-1. Comparative Overall Test Method Accuracy Characteristics for IRE[1], ICE[2], HET-CAM[3], and BCOP[4] in Identifying GHS[5] Ocular Corrosives/ Severe Irritants (UN[6] [2003]) – Reanalyses**

| Statistic | IRE | | | ICE | | HET-CAM | | | BCOP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Old[7] (n = 36)[8] | New[7] (n = 38) | Expanded-New[9] (n = 76) | Old (n = 92) | New (n = 144) | Old (n = 52) | New[10] (n = 101) | New[11] (n = 143) | Old (n = 120) | New (n = 147) |
| **Accuracy** | 78% (28/36) | 79% (30/38) | 68% (52/76) | 82% (75/92) | 83% (120/144) | 85% (44/52) | 68% (69/101) | 53% (76/143) | 79% (95/120) | 81% (119/147) |
| **Sensitivity** | 100% (12/12) | 100% (11/11) | 100% (33/33) | 60% (15/25) | 50% (15/30) | 100% (12/12) | 70% (28/40) | 85% (35/41) | 76% (32/42) | 84% (36/43) |
| **Specificity** | 67% (16/24) | 70% (19/27) | 44% (19/43) | 90% (60/67) | 92% (105/114) | 80% (32/40) | 67% (41/61) | 40% (41/102) | 81% (63/78) | 80% (83/104) |
| **Positive Predictivity** | 60% (12/20) | 58% (11/19) | 58% (33/57) | 68% (15/22) | 63% (15/24) | 60% (12/20) | 58% (28/48) | 36% (35/96) | 69% (34/49) | 63% (36/57) |
| **Negative Predictivity** | 100% (16/16) | 100 (19/19) | 100% (19/19) | 86% (60/70) | 88% (105/120) | 100% (32/32) | 77% (41/53) | 87% (41/47) | 86% (61/71) | 92% (83/90) |
| **False Positive Rate** | 33% (8/24) | 30% (8/27) | 56% (24/43) | 10% (7/67) | 8% (9/114) | 20% (8/40) | 33% (20/41) | 60% (61/102) | 19% (15/78) | 20% (21/104) |
| **False Negative Rate** | 0% (0/12) | 0% (0/11) | 0% (0/33) | 40% (10/25) | 50% (15/30) | 0% (0/12) | 30% (12/40) | 15% (6/35) | 24% (10/42) | 16% (7/43) |

[1]IRE = Isolated Rabbit Eye assay.
[2]ICE = Isolated Chicken Eye assay.
[3]HET-CAM = Hen's Egg Test – Chorioallantoic Membrane assay.
[4]BCOP = Bovine Corneal Opacity and Permeability assay.
[5]GHS = Globally Harmonized System.
[6]UN = United Nations.
[7]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the analysis included in the IRE draft BRD with corrections.
[8]n = number of substances tested; the numbers in parentheses in each row indicates the data on which the percentage calculation is based.
[9]Includes the 38 substances tested by Guerriero et al. (2004) and 38 unique substances classified as severe irritants in Balls et al. (1995) and Gettings et al. (1996), based either on an *in vitro* corneal opacity score of at least 3.0 or an *in vitro* corneal swelling of at least 25%; these were among the criteria used by Guerriero et al. (2004) to identify corrosive/severe irritants.
[10]These data are for the IS(B) method (described by Kalweit et al. 1987) when testing substances as a 10% solution *in vitro*.
[11]These data are for the IS(B) method (described by Kalweit et al. 1987) when testing substances at a 100% concentration *in vitro*.

A reanalysis of the accuracy of the IRE test method for identifying ocular corrosives and severe irritants based on the reclassification of some nonsevere irritants as severe irritants was conducted. The results are independent of the three classification systems used; thus the discussion here is limited to the GHS classification system. When the reanalysis is restricted to Guerriero et al. (2004), the accuracy[1] changed from 78% (28/36) in the draft IRE BRD to 79% (30/38) in the reanalysis, the false negative rate stayed the same (draft IRE BRD = 0% [0/12]; reanalysis: 0% [0/11]) and the false positive rate decreased from 33% (8/24) in the draft IRE BRD to 30% (8/27) in the reanalysis.

For the expanded data set and using the GHS ocular hazard classification system, the accuracy was 68% (52/76), the false negative rate was 0% (0/33), and the false positive rate was 56% (24/43). The expanded data set used for this evaluation include the 38 substances evaluated by Guerriero et al. (2004) and an additional 38 substances tested by Balls et al. (1995) and Gettings et al. (1996) and classified by IRE as severe irritants, 22 of which were also severe irritants *in vivo* and 16 of which were nonsevere irritants or nonirritants *in vivo*. The expanded data set is potentially confounded by the exclusion of substances with true negative outcomes (matching *in vivo* and *in vitro* nonsevere or nonirritant classifications), which would affect both specificity and the false positive rate.

In order to further evaluate discordant responses of the IRE test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form). Because the international community will soon adopt the GHS classification system for hazard labeling (UN [2003]), and considering that there were only modest differences in overall IRE test method accuracy among the three regulatory classification systems (i.e., EPA, EU, GHS), these sub-analyses are focused only on the GHS classification system, using the expanded data set.

The chemical classes that had the highest rate of IRE test method overprediction according the GHS classification system (i.e., were false positives) were ketones (67%, [4/6]), esters (67%, [4/6]), and alcohols (60%, [6/10]). Among the 10 surfactants tested, the false positive rate was 50% (2/4) and the false negative rate was 0% (0/6). The seven cationic surfactants included in this group had a false positive rate of 100% (1/1) and a false negative rate of 0% (0/6).

[1] Accuracy is defined as the proportion of correct outcomes (positive and negative) of a test method; Sensitivity is defined as the proportion of all positive substances that are classified as positive; Specificity is defined as the proportion of all negative substances that are classified as negative; Positive predictivity is defined as the proportion of correct positive responses among substances testing positive; Negative predictivity is defined as the proportion of correct negative responses among substances testing negative; False positive rate is defined as the proportion of all negative substances that are falsely identified as positive; False negative rate is the defined as the proportion of all positive substances that are falsely identified as negative (ICCVAM 1997).

With regard to physical form of the substances overpredicted by the IRE test method, liquids had a higher overprediction rate (83%, [19/23]) than solids (25%, [5/20]). There was insufficient data to analyze the effect of pH on overprediction.

No substances were underpredicted (i.e., were false negatives) by the IRE test method. Thus, an analysis of underprediction based on chemical class, physical form, pH, or NICEATM GHS Category I subclassification was not possible.

In the original draft IRE BRD (NICEATM [2004]), no data was provided for the assessment of intralaboratory repeatability and reproducibility. Since no additional data was submitted for the IRE test method following the Expert Panel meeting, an analysis of intralaboratory reliability still could not be conducted.

The original IRE test method reliability analysis included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the reclassification of the ocular irritancy of some test substances, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/non-severe irritant) of the individual laboratory test results obtained for the EC/HO validation study (Balls et al., [1995]) and for the CEC (1991) collaborative study was affected.

Overall, in the Balls et al. (1995) study, the number of substances with 100% agreement among the four participating laboratories was 59-61% (35-36/59) in the original analysis and 59-63% (35-37/59) in the reanalysis. The number of substances with 75% agreement among laboratories was 24-25% (14-15/59) in the original analysis and 22-25% (13-15/59) in the reanalysis. The number of substances with 50% agreement among laboratories did not change due to the reanalysis (15% [9/59 substances]).

Overall, in the CEC (1991) study, the number of substances with 100% agreement among the three participating laboratories decreased from 86% (18/21) to 81% (17/21) in the reanalysis. The number of substances with 67% agreement among laboratories remained the same at 14% (3/21), while the number of substances with 33% agreement was increased from 0% to 5% (1/21).

***ICE Test Method***

The ICE test method protocol (also referred to as the Chicken Enucleated Eye Test [CEET]) was first described by Prinsen and Koëter (1993) and was developed based on the IRE test developed by Burton et al. (1981). In this *in vitro* bioassay, the test substance is applied to the cornea of eyes isolated from chickens that have been slaughtered for human consumption. Three parameters are evaluated to measure the extent of damage to the eye following exposure to a chemical substance: corneal swelling, corneal opacity, and fluorescein retention. While the latter two parameters involve a subjective assessment, analysis of corneal swelling provides an objective measurement, thus potentially providing improved precision and reduced interlaboratory variability compared to the traditional *in vivo* rabbit eye test, which relies only on subjective measurements.

For this reanalysis, additional ICE test method data and corresponding *in vivo* rabbit eye test data were submitted by the TNO Nutrition and Food Institute for the 44 substances tested in Prinsen (1996) and for an additional 50 substances (Prinsen [2005]). Also, the TNO Nutrition and Food Institute provided replicate ICE test data and the corresponding *in vivo* EU hazard classification for four substances (Prinsen [2000]). The additional data increased the number of substances in the comparative ICE:*in vivo* rabbit eye test database from 92 to 149 substances for the GHS classification system (UN [2003]), from 90 to 148 for the U.S. Environmental Protection Agency (EPA) classification system (EPA [1996]), and from 121 to 155 for the EU classification system (EU [2001]).

Depending on the classification system used, the overall accuracy of the ICE test method changed from 82-83% (old analysis) to 83-84% (reanalysis), the false positive rate was reduced from 8-10% (old analysis) to 6-8% (reanalysis), while the false negative rate was increased from 30-40% (old analysis) to 40-50% (reanalysis).

Consistent with the original analysis, the reanalysis indicated that alcohols are overpredicted (50% [5/10] false positive rate) in the ICE test method. Carboxylic acids were shown to have a false negative rate of 43% (3/7).

The total database for surfactants was increased from 13 to 21 substances. However, given the stability of the false negative rate (old analysis: 57% [4/7]; new analysis 56% [5/9]), these substances still appear to be underpredicted by the ICE test method. With the additional data, it was now possible to evaluate the accuracy of the ICE test method for pesticides. While the false positive rate for these substances was 0% (0/6), the false negative rate for pesticides was 60% (3/5).

Eight of the fifteen underpredicted substances were liquids while seven were solids. However, considering that the total number of solids (36) in the database is much smaller than the number of liquids (108), solids appear more likely to be underpredicted (58%) than liquids (44%) by the ICE test method. In comparison to the original analysis, the false negative rate of solid substances increased from 55% (6/11) to 58% (7/12), while that for liquids increased from 29% (4/14) to 44% (8/18).

Using the expanded database, an analysis was conducted of the ability of the ICE test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant. Underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions only (false negative rate = 70% [7/10]), than on severe lesions (false negative rate = 45% [9/20]).

A new analysis not included originally was an evaluation of accuracy related to acidic or basic pH. Among the eight underpredicted substances for which pH information was available, four were acidic (pH < 7.0) and four were basic (pH > 7.0). Again, basic substances (8) occupy a smaller proportion of the total database than acidic substances (12),

and were more often underpredicted (50% vs. 33%). However, pH information was obtained for only 20 of the 30 total Category 1 substances.

Previously, an evaluation of the intralaboratory repeatability and reproducibility of the ICE test method could not be conducted. However, subsequent to the original reliability analysis, data were received that allowed for a quantitative analysis of intralaboratory repeatability and reproducibility of ICE test method endpoints.

The range of percent coefficient of variation (%CV) values for the corneal thickness measurement, when results were compared within experiments, was from 0.9% to 6.1%. The other endpoints evaluated produced ranges of %CV values that were larger, with variability most prominent with the nonirritating substance (SP-1). However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher %CV than values of 11, 14, and 18). A similar discussion can also be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively).

The range of %CV values for the corneal thickness measurement, when results were compared across laboratories, was from 1.8% to 6.3%. The %CV values for the remaining endpoints had a larger range (e.g., corneal swelling %CV = 13.9% to 138.7%). However, if the nonirritating substance is removed, the range of %CV values is reduced (e.g., corneal swelling %CV = 13.9% to 22.4%).

The previous analysis also included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the new information that was received, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/non-severe irritant) of the individual laboratory test results obtained for the EC/HO validation study (Balls et al., [1995]) needed to be repeated. However, the results obtained in the revised analysis were not different from the original analysis.

***BCOP Test Method***

The BCOP assay is an *in vitro* eye irritation test method using isolated bovine eyes from cattle that have been slaughtered for meat or other purposes. In the BCOP assay, opacity is determined by the amount of light transmission through the cornea, and permeability is determined by the amount of sodium fluorescein dye that passes through all corneal cell layers. More recent additions/endpoints to the BCOP assay are assessment of corneal swelling or hydration, and histological assessment of morphological alterations in the cornea (Bruner et al. [1998]; Ubels et al. [1998]; Cooper et al. [2001]; Jones et al. [2001]). When histological assessment is added to the BCOP assay, the type and depth of corneal injury can be evaluated, as well as whether the tissue damage is permanent (e.g., damage to the endothelium) (Gran et al. [2003]).

Subsequent to the draft BCOP BRD, *in vivo* rabbit eye test data that corresponded to the substances tested in BCOP in the Gautheron et al. (1994) study were received from Johnson & Johnson Pharmaceutical R&D. Individual cornea data from the BCOP tests evaluating these 52 substances were also provided subsequent to the meeting. Johnson & Johnson Pharmaceutical R&D also provided individual cornea data for 20 substances evaluated in the BCOP test method, comparing results achieved using corneas from adult animals (>24 months) versus those from young animals (6 - 8 months). The additional data increased the size of the comparative BCOP:*in vivo* rabbit eye test database from 120 to 147 substances for the GHS classification system (UN [2003]), 117 to 143 for the EPA classification system (EPA [1996]). In contrast, due to changes in study acceptability criteria (i.e., the classification call needed to be based on *in vivo* rabbit eye test data), the size of the comparative BCOP:*in vivo* rabbit eye test database was decreased from 157 to 143 substances for the EU classification system (EU [2001]).

The overall accuracy stayed the same in the reanalysis evaluation (original analysis: 77-80%, depending on the classification system used; reanalysis: 80% for all classification systems). The false positive rate was reduced from 23% (original analysis) to 21% (reanalysis) for the EU classification system (EU [2001]), but was increased from 17-19% (original analysis) to 19-20% (reanalysis) for the EPA (EPA [1996]) and GHS (UN [2003]) classification systems, respectively; while the false negative rate was reduced for all three classification systems (from 23-27% [original analysis] to 16-25% [reanalysis]).

Similar to the original analysis, the reanalysis indicated that alcohols are often overpredicted (50% [9/18] false positive rate) in the BCOP test method. Carboxylic acids (3/9) and heterocyclic compounds (2/6) had a false negative rate of 33%.

Eighteen of the 20 overpredicted substances were liquids while two were solids. Considering the proportion of the total available database, liquids (93) appear more likely than solids (34) to be overpredicted by the BCOP test method. In comparison to the original analysis, the overprediction of solid substances was reduced (from 44% [4/9] to 10% [2/20] false positive rate), while the false positive rate for liquids was increased from 21% (14/66) to 26% (18/69).

With regard to physical form of the substances underpredicted by the BCOP test method, six were solids and one was a liquid. In comparison to the original analysis, the false negative rate for solid substances was increased from 31% (4/13) to 43% (6/14), while the false negative rate for liquids was reduced in the revised analysis from 18% (5/28) to 4% (1/24).

Using the expanded database, an analysis was conducted of the ability of the BCOP test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant. The underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (false negative rate = 23% [3/13]), rather than on severe lesions (false negative rate = 17% [4/24]).

A new analysis not included originally was an evaluation of accuracy related to acidic or basic pH. Among the five underpredicted substances for which pH information was available, two (18% [2/11]) were acidic (pH < 7.0) and three (23% [3/13]) were basic (pH > 7.0). pH information was obtained for only 24 of the 43 total Category 1 substances.

The analyses of intralaboratory reliability were not affected by the information received subsequent to the release of the draft BCOP BRD. However, the previous analysis also included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the new information that was received, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/nonsevere irritant) of the data provided for multiple laboratories in three studies (Gautheron et al. [1994]; Balls et al. [1995]; Southee [1998]) needed to be repeated.

The results obtained in the revised analysis of interlaboratory reproducibility were not different from the original analysis. The five participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested *in vitro* in the study. In general, the extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% to 86%, depending on the classification system used, of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), there was 100% agreement in regard to the ocular irritancy classification for 35 to 36 (67% to 69%) of the 52 substances, which were tested in either 11 or 12 laboratories. Finally, for the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances.

***HET-CAM Test Method***

The HET-CAM test method uses the chorioallantoic membrane (CAM), which is a vascular fetal membrane composed of the fused chorion and allantois. The method is proposed to provide information on the effects that may occur in the conjunctiva following exposure to a test substance. It was assumed that acute effects induced by a test substance on the small blood vessels and proteins of this soft tissue membrane are similar to effects induced by the same test substance in the eye of a treated rabbit. The CAM has been proposed as a model for a living membrane (such as the conjunctiva) since it comprises a functional vasculature. Additionally, evaluation of coagulation (i.e., protein denaturation) may reflect corneal damage that may be produced by the test substance. The CAM is evaluated for the development of irritant endpoints (hyperemia, hemorrhage, and coagulation). Depending on the method used to collect data on the endpoints (time to development, severity of observed effect), qualitative assessments of the irritation potential of test substances are made.

Additional HET-CAM test method data and corresponding *in vivo* rabbit eye test data were received from ZEBET for substances that were originally described in Spielmann et al. (1996) (Spielmann and Liebsch [2005a]). HET-CAM test data previously discussed in

Section 9.0 of the draft HET-CAM BRD also were included in this reanalysis (Gilleron et al. [1996, 1997]). Results from control studies run concurrently with HET-CAM studies also were provided (Vanparys and VanGoethem [2005b]; Spielmann and Liebsch [2005b]). In addition, replicate intralaboratory and interlaboratory HET-CAM test data were obtained (Vanparys and VanGoethem [2005a]).

When the reanalyses were conducted with the IS(A) and IS(B) methods[2], based on the additional data received, wherein substances tested at either 10% or 100% concentration were compared only against *in vivo* studies which had been conducted with undiluted test substances, the following patterns were noted. For the IS(A) analysis method, test method accuracy increased when substances were evaluated at 100% concentration *in vitro* compared to the 10% concentration (e.g., 85% [17/20] for IS(A)-100 vs. 50% [12/24] for IS(A)-10; GHS classification system). In comparison, the opposite pattern was observed for the IS(B) analysis method; test method accuracy increased when substances were evaluated at 10% concentration (IS(B)-10) *in vitro* compared to the 100% concentration (IS(B)-100) (e.g., 68% [69/101] for IS(B)-10 vs. 53% [76/143] for IS(B)-100; GHS classification system).

Unlike the draft HET-CAM BRD analysis, where only formulations were evaluated by the IS(B) method, additional chemical classes were available for this reassessment. The revised analysis indicated that there are several chemical classes that are overpredicted by the HET-CAM IS(B) analysis methods when testing substances at either a 10% or at 100%. These chemical classes include alcohols (IS(B)-10: 90% [9/10]; IS(B)-100: 91% [10/11]), ethers (IS(B)-10: 50% [5/10]; IS(B)-100: 60% [9/15]), amines (IS(B)-10: 60% [3/5]; IS(B)-100: 83% [5/6]), organic salts (IS(B)-10: 57% [4/7]; IS(B)-100: 88% [7/8]), and heterocycles (IS(B)-10: 83% [5/6]; IS(B)-100: 75% [6/8]). Additionally, the IS(B)-100 analysis method overpredicted esters (83% [10/12]). The chemical class that was consistently underpredicted by the analysis methods was formulations (IS(B)-10: 44% [7/16]; IS(B)-100: 35% [7/13]).

An evaluation based on the physical form of the test substance depended on the analysis method being evaluated. Liquids could only be evaluated for the IS(B)-10 analysis method while solids and liquids could be evaluated for the IS(B)-100 analysis method. In the case of the IS(B)-100 evaluation, solids had a higher false positive rate when compared to liquids (76% [16/21] vs. 60% [36/60]). In contrast, the false negative rates for solids and liquids were approximately equal (IS(B)-10: 30%, 10/33 for liquids; IS(B)-100: 28% [7/25] and 26% [5/19] for solids). The false positive and false negative rate for liquids (when tested by the IS(B)-10 method) also were approximately equal (false positive: 34% [21/62]; false negative: 30% [10/33]).

An analysis of the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant,

[2] IS(A) analysis method refers to the method of Luepke (1985). This method evaluates the development of endpoints at pre-determined time points (e.g., 0.5, 2 and 5 minutes) and assigns a score based on the time of appearance of endpoint. The scores are totaled to determine an irritation score. IS(B) analysis method refers to the method Kalweit et al. (1987). This method determines the time required for endpoints to develop and uses these values to develop an irritation score.

indicated that, for IS(B)-10, the underpredicted substances were more likely to be substances classified as corrosive or severely irritating *in vivo* based on persistent lesions, with a false negative rate of 37% (10/27) compared to 15% (2/13) for substances classified as corrosive or severely irritating *in vivo* based on severity. This was not true for IS(B)-100, where the false negative rates for both persistent and severely irritating substances were the same (11% [2/18] and 11% [2/19], respectively).

Previously, an evaluation of the intralaboratory repeatability and reproducibility of the HET-CAM test method could not be conducted. However, subsequent to the release of the draft HET-CAM BRD, replicate within and among test data were received that allowed for a quantitative analysis of intralaboratory repeatability and reproducibility of HET-CAM test method endpoints.

The analysis of intralaboratory repeatability (i.e., the extent of variability among replicate eggs in the same study) was evaluated using data from two different publications (Gilleron et al. [1996, 1997]) that were provided by the authors in response to a request from NICEATM. In both studies, the highest %CV values were associated with the hemorrhage endpoint (104-117%), while the lowest %CV values were associated with the measurement of coagulation (38%-115%). However, the actual values were quite disparate between the two studies (e.g., Gilleron et al. [1996] coagulation %CV = 115.07, Gilleron et al. [1997] coagulation %CV = 37.78). The difference in the numbers between the two studies may be due to several factors including the nature of the test substances evaluated and differences in the test method protocols used. The mean and median overall *in vitro* score %CV for all substances tested was 32.52% for Gilleron et al. (1996) and 7.61 for Gilleron et al. (1997). The calculated intralaboratory repeatability for the endpoints and the overall test method may be exaggerated because of the relatively small values that are obtained from each of the endpoints (from 0 to 5 for hemorrhage, 0 to 7 for lysis, and 0 to 9 for coagulation).

Similar results were obtained from the analysis of intralaboratory reproducibility. The overall %CV values were 53.0% and 17.5% for the two studies evaluated. For the study by Gilleron et al. (1997), where substances could be classified according to the GHS and EPA classification systems, %CV values for severe irritants were similar to the values obtained for the overall database.

The previous analysis also included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. Additional data received subsequent to the draft HET-CAM BRD allowed for a more in-depth quantitative and qualitative analysis of interlaboratory reproducibility. A qualitative evaluation of data from Spielmann et al. (1996) indicates that that the level of agreement in classification of a test substance between testing laboratories, when evaluated per the GHS classification system, is 79% (85/107) and 82% (81/99) for the IS(B)-10 and IS(B)-100 analysis methods, respectively. A quantitative evaluation of the interlaboratory reproducibility of the test method based on a %CV analysis resulted in a mean %CV values of 60.17 for the IS(B)-10 analysis method and 35.21 for the IS(B)-100 analysis method.

The previous interlaboratory reproducibility analyses also were modified based on the reclassification of substances as an ocular corrosive/severe irritant or as a noncorrosive/ nonsevere irritant. However, the overall results obtained in the revised analysis were not different from the original analysis.

Finally, historical positive and negative control data were provided by two different sources. The negative control substance evaluated was 0.9% sodium chloride. The positive control substances were dimethylformamide, imidazole, 1% sodium dodecyl sulfate, and 0.1 N sodium hydroxide. The studies showed that all control substances consistently produced appropriate responses (e.g., negative control consistently produced a response that would be classified as nonirritant and positive controls consistently produced a response that would be classified as severe irritant).

***Reference Substances***

Included in each draft BRD [NICEATM 2004] were a list of proposed reference substances for the optimization and/or validation of *in vitro* tests to detect ocular corrosives and severe irritants (available electronically at http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm). The proposed reference substances are intended to:

- represent the range of ocular responses (i.e., corrosive/severe irritant; nonsevere irritant/noncorrosive) that the test method is expected to be capable of predicting
- represent the range of chemical/product classes and physicochemical properties (e.g., solid, liquid) that the test method is expected to be capable of testing
- represent the range of known or anticipated mechanisms or modes of action for severe/irreversible ocular irritation or corrosion
- have been generated by high-quality *in vivo* rabbit eye test method studies following Organization for Economic Cooperation and Development (OECD) Test Guideline (TG) 405 (OECD [1987]) and preferably conducted in compliance with Good Laboratory Practices (GLP) guidelines (OECD [1998]; EPA [2004a, 2004b]; FDA [2004])
- have a well-defined chemical composition
- be tested at a defined concentration and at a defined purity[3]
- be readily available

[3]Information on purity and the concentration tested were not available for all substances included in the NICEATM *in vivo* rabbit eye test results database. A decision was made to exclude nonsevere irritants (i.e., GHS Category 2A or 2B irritants) or non-irritants but not corrosive/severe irritants (i.e., GHS Category 1) that lacked concentration data from consideration as proposed reference substances. GHS category 1 substances were included because testing at a potentially higher concentration would not likely alter their classification as a GHS Category 1 substance although it might alter the criteria by which they were classified as an ocular corrosive/severe irritant. Where information on purity was lacking, an assumption was made that testing would have been conducted with a relatively pure substance. For substances included because they cause severe ocular effects in humans but lacked appropriate *in vivo* rabbit eye test data, information on concentration and purity were not available.

The Expert Panel concluded that the list of proposed substances is fairly comprehensive in that the three major groups of products to which the eye is exposed (i.e., industrial chemicals, pharmaceuticals, cosmetics) are represented and that, in general, individual substances were appropriately chosen. The Expert Panel also suggested several changes to the list of proposed reference substances. In response to their recommendations, a revised list of proposed reference substances has been developed. This list includes 11 more inorganic substances (especially those used in consumer products) and ten substances that are known human ocular corrosives or severe irritants (even in the absence of high quality Draize rabbit eye test data), contains fewer surfactants, and excludes formulations. In contrast, all 12 formulations in the original proposed list have been excluded, and the number of surfactants has been reduced from 12 to seven. In addition,

- the source of the Draize rabbit eye test data has been provided for each proposed reference substance
- where applicable and to the extent possible, within a chemical class, substances of lower, medium and higher molecular weight have been included (the molecular weight of each proposed substance is now provided)
- information is provided on whether each proposed reference substance has been tested in the proposed version of BCOP, HET-CAM, ICE, and IRE test methods

In addition to considering the recommendations of the Expert Panel, clarification regarding the rules for classification of severe irritants was obtained subsequent to the release of the four BRDs that resulted in changes to the hazard classification of a few of the substances included in the original list of proposed reference substances. Also, the chemical classes assigned to each reference substance were revised to be consistent with MeSH, an internationally recognized standardized classification scheme (Available: http://www.nlm.nih.gov/mesh [NLM 2005]). Finally, additional Draize rabbit eye test results for about 300 substances were obtained from several sources that expanded the number of potential candidate substances for consideration.

The revised list contains 122 substances including 79 GHS Category 1 substances (10 of which were classified as severe irritants based on human data only), 28 GHS Category 2 substances (14 Category 2A, 13 GHS Category 2B, 1 GHS Category 2A/2B) and 15 noninrritants. For the detection of ocular corrosives and severe irritants, the list of reference substances needs to include substances that:

- induce very severe responses within a relatively short time period, as well as those where the toxic response is delayed
- adversely affect the cornea, iris, and/or conjunctiva
- induce persistent versus non-persistent lesions (when assessed at 21 days post treatment)
- represent diverse chemical classes and physicochemical properties

The total number of proposed reference substances reflects the additional substances recommended by the Expert Panel and the need to ensure, to the extent possible, that the substances covered the range of responses of interest, chemical/product classes and

physicochemical properties of interest, and known or anticipated mechanisms or modes of action for severe/irreversible ocular irritation or corrosion. Nevertheless, power calculations are being conducted by NICEATM to evaluate the appropriateness of this number of substances for evaluating the accuracy of an *in vitro* ocular irritancy test method.

This list of proposed reference substances is intended to represent the minimum number of substances considered critical to an evaluation of the validity of alternative *in vitro* ocular irritancy test methods, while subsets of substances from this list may be considered for:

- optimization of a test method protocol
- performance standard reference substances for use in the validation of test methods that are functionally and mechanistically similar to a validated ocular irritancy test method
- proficiency testing to ensure the competency of a laboratory in performing a validated ocular irritancy test method

## References

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.

Bruner LH, Evans MG, McPherson JP, Southee JA, Williamson PS. 1998. Investigation of ingredient interactions in cosmetic formulations using isolated bovine corneas. Toxicol In Vitro 12:669-690.

Burton ABG, York M, Lawrence RS. 1981. The *in vitro* assessment of severe eye irritants. Food Cosmet Toxicol 19:471-480.

CEC. 1991. Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. Doc. XI/632/91/V/E/1/131/91 Part I and II.

Cooper KJ, Earl LK, Harbell J, Raabe H. 2001. Prediction of ocular irritancy of prototype shampoo formulations by the isolated rabbit eye (IRE) test and bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 15:95-103.

EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

EPA. 2004a. Good laboratory practice standards. Toxic Substances Control Act. 40 CFR 792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_04/40cfr792_04.html [accessed 12 July 2005]

EPA. 2004b. Good laboratory practice standards. 40CFR160. Available: http://www.access.gpo.gov/nara/cfr/waisidx_04/40cfr160_04.html [accessed 12 July 2005].

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws,

regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L255:1-333.

FDA. 2004. Good laboratory practice for nonclinical laboratory studies. 21CFR58. Available. http://www.access.gpo.gov/nara/cfr/waisidx_04/21cfr58_04.html [accessed 12 July 2005].

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Tos EG, Hagemann C, Vanparys P, Deknudt G, Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 8:381-392.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, Demetrulias JL, Dipasquale LC, Earl LK, Feder PI, Galli CL, Glaza SM, Gordon VC, Janus J, Kurtz PJ, Marenus KD, Moral J, Pape WJW, Renskers KJ, Rheins LA, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J. 1996. The CTFA evaluation of alternatives program: An evaluation of *in vivo* alternatives to the Draize primary rabbit eye irritation test. (Phase III) Surfactant-based formulations. Food Chem Toxicol 34:79-117.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1997. Evaluation of the HET-CAM-TSA method as an alternative to the Draize eye irritation test. Toxic In Vitro 11:641-644.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1996. Evaluation of a modified HET-CAM assay as a screening test for eye irritancy. Toxic In Vitro 10:431-446.

Gran BP, Swanson JE, Merill JC, Harbell JW. 2003. Evaluating the irritancy potential of sodium percarbonate: a case study using the bovine corneal opacity and permeability (BCOP) assay. The Toxicologist 72:220.

Guerriero FJ, Seaman CW, Olson MJ, Guest R, Whittingham A. 2004. Retrospective assessment of the rabbit enucleated eye test (REET) as a screen to refine worker safety studies. [Abstract No. 1282] Toxicol Sci (The Toxicologist Supplement) 78(1-S).

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm [accessed 12 July 2005].

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Methods: A Report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No: 97-3981. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate/pdf [accessed 12 July 2005].

Jones PA, Budynsky E, Cooper KJ, Decker D, Griffiths HA, Fentem JH. 2001. Comparative evaluation of five *in vitro* tests for assessing the eye irritation potential of hair-care products. ATLA 29:669-692.

Kalweit S, Gerner I, Spielmann H. 1987. Validation project of alternatives for the Draize eye test. Mol Toxic 1:597-603.

Luepke N. 1985. Hen's egg chorioallantoic membrane test for irritation potential. Food Chem Toxic 23:287-291.

NICEATM. 2004. Draft Background Review Document. Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Isolated Rabbit Eye (IRE), the Isolated Chicken Eye (ICE), the Bovine Corneal Opacity and Permeability, and the Hen's Egg Test – Chorioallantoic Membrane Test Methods. Available: http://iccvam.niehs.nih.giv/methods/ocudocs/ocu_brd.htm [accessed 15 July 2005].

NICEATM. 2005a. Evaluation of the Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 15 July 2005].

NICEATM. 2005b. *In Vitro* Ocular Test Method Accuracy and Reliability Reanalyses. (Electronic form available: http://iccvam.niehs.nih.gov/methods/ocudocs/reanalysis.htm [accessed 15 July, 2005]; Paper form available: niceatm@niehs.nih.gov.

NIEHS. 2005. Second request for data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. Fed Reg 70:9661-9662. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 15 July 2005].

NLM. 2005. U.S. National Library of Medicine's Medical Subject Headings. Available: http://www.nlm.nih.gov/mesh [accessed 15 July 2005].

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD principles on Good Laboratory Practice. (as revised in 1997). ENV/MC/CHEM(98)17. Paris: OECD. Available: http://www.oecd.org [accessed 12 July 2005].

OECD. 2002. Test Guideline 405. Acute eye irritation/corrosion, adopted April 24, 2002. In: OECD Guidelines for Testing of Chemicals. Paris: OECD.

Prinsen MK. 2005. *In vitro* and *in vivo* data for 94 substances tested in the isolated chicken eye test. Unpublished data provided directly to NICEATM by MK Prinsen, TNO Nutrition and Food Research Institute.

Prinsen MK. 2000. Chicken enucleated eye test with reference surfactants and siloxane polymers in Phase II of the reference standard validation project; an *ex vivo* alternative to the

Draize eye irritation test with rabbits. TNO Report V99.521b. Unpublished report provided directly to NICEATM by M Prinsen, TNO Nutrition and Food Research Institute.

Prinsen MK. 1996. The chicken enucleated eye test (CEET): A practical (pre)screen for the assessment of eye irritation/corrosion potential of test materials. Food Chem Toxicol 34:291-296.

Prinsen MK, Koëter BWM. 1993. Justification of the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. Food Chem Toxicol 31:69-76.

Southee JA. 1998. Evaluation of the Prevalidation Process. Part 2, final report. Volume 2. The Bovine Corneal Opacity and Permeability (BCOP) Assay. European Community contract no. 11279-95-10F 1ED ISP GB.

Spielmann H, Liebsch M. 2005a. In vivo data for substances used to evaluate the accuracy of the HET-CAM test method. Unpublished data provided directly to NICEATM by H. Spielmann and M. Liebsch, ZEBET.

Spielmann H, Liebsch M. 2005b. . In vitro reproducibility and repeatability data for control chemicals tested in the HET-CAM test method. Unpublished data provided directly to NICEATM by H. Spielmann and M. Liebsch, ZEBET.

Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhütter H, Schneider B, Glaser S, Gerner I, Pape WJW, Kreiling R, Krauser K, Miltenburger HG, Steiling W, Luepke NP Müller N, Kreuzer H, Mürmann P, Spengler J, Bertram-Neis, E, Siegemund B, Wiebel F. 1996. Results of a validation study in Germany on two in vitro alternatives to the Draize eye irritation test, HET-CAM test and the 3T3 NRU cytotoxicity test. ATLA 24:741-858.

Ubels JL, Erickson AM, Zylstra U, Kreulen CD, Casterton PL. 1998. Effect of hydration on opacity in the bovine corneal opacity and permeability (BCOP) assay. J Toxicol – Cut and Ocular Toxicol 17(4):197-220.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http:/www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html [accessed 15 July, 2005].

Vanparys P, VanGoethem F. 2005a. In vitro reproducibility and repeatability data for 106 chemicals tested in the HET-CAM test method. Unpublished data provided directly to NICEATM by P. Vanparys and F. VanGoethem, Johnson & Johnson.

Vanparys P, VanGoethem F. 2005b. In vitro reproducibility and repeatability data for control chemicals tested in the HET-CAM test method. Unpublished data provided directly to NICEATM by P. Vanparys and F. VanGoethem, Johnson & Johnson.

***[This Page Intentionally Left Blank]***